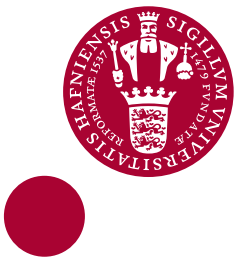

The cytochrome P450 CYP72A552 is key to production of hederageninbased saponins that mediate plant defense against herbivores

Liu, Qing; Khakimov, Bekzod; Cárdenas, Pablo D.; Cozzi, Federico; Olsen, Carl Erik; Jensen, Karen Rysbjerg; Hauser, Thure Pavlo; Bak, Søren

Published in:
New Phytologist (Online)

DOI:
10.1111/nph.15689

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

New Phytologist

Research

The cytochrome P450 CYP72A552 is key to production of hederagenin-based saponins that mediate plant defense against herbivores

Qing Liu[1], Bekzod Khakimov[1,2], Pablo D. Cárdenas[1], Federico Cozzi[1,3], Carl Erik Olsen[1], Karen Rysbjerg Jensen[1], Thure Pavlo Hauser[1] and Søren Bak[1]

[1]Department of Plant and Environmental Sciences and Copenhagen Plant Science Center, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark; [2]Department of Food Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark; [3]BIOMIN Research Center, Technopark 1, 3430 Tulln an der Donau, Austria

Authors for correspondence:
Søren Bak
Tel: +45 35333346
Email: bak@plen.ku.dk

Qing Liu
Tel: +45 35331495
Email: qliu@plen.ku.dk

Received: *29 October 2018*
Accepted: *2 January 2019*

New Phytologist (2019) **222:** 1599–1609
doi: 10.1111/nph.15689

Key words: cytochrome P450, molecular evolution, pathway elucidation, plant–insect interaction, triterpenoid saponin biosynthesis.

Summary

- Plants continuously evolve new defense compounds. One class of such compounds is triterpenoid saponins. A few species in the *Barbarea* genus produce saponins as the only ones in the large crucifer family. However, the molecular mechanism behind saponin biosynthesis and their role in plant defense remains unclear.
- We used pathway reconstitution *in planta*, enzymatic production of saponins *in vitro*, insect feeding assays, and bioinformatics to identify a missing gene involved in saponin biosynthesis and saponin-based herbivore defense.
- A tandem repeat of eight *CYP72A* cytochromes P450 colocalise with a quantitative trait locus (QTL) for saponin accumulation and flea beetle resistance in *Barbarea vulgaris*. We found that CYP72A552 oxidises oleanolic acid at position C-23 to hederagenin. *In vitro*-produced hederagenin monoglucosides reduced larval feeding by up to 90% and caused 75% larval mortality of the major crucifer pest diamondback moth and the tobacco hornworm. Sequence analysis indicated that *CYP72A552* evolved through gene duplication and has been under strong selection pressure.
- In conclusion, CYP72A552 has evolved to catalyse the formation of hederagenin-based saponins that mediate plant defense against herbivores. Our study highlights the evolution of chemical novelties by gene duplication and selection for enzyme innovations, and the importance of chemical modification in plant defense evolution.

Introduction

Plants continuously evolve new defense compounds to survive in the 'arms race' against herbivores. Saponins are one such class of specialised compounds. They consist of a hydrophobic triterpenoid backbone with one or more hydrophilic saccharide groups attached. Some saponins are toxic or deterrent to insects, molluscs, fungi and other microorganisms, and have probably evolved for defense against these, although their mode of action and evolution is not fully understood (De Geyter *et al.*, 2007). Research on different Lepidoptera (moths and butterflies) cell lines indicates that saponins can affect midgut-cell membranes (De Geyter *et al.*, 2012), which may explain their toxicity. In addition to cell membrane disruption, saponins can also act as deterrents to insect feeding due to their bitter taste (De Geyter *et al.*, 2007; Augustin *et al.*, 2012).

The ability to produce saponins has evolved convergently in several different plant families, testifying their biological significance. Within the large crucifer family (Brassicaceae), only a few species in the genus *Barbarea* are known to produce saponins (Badenes-Perez *et al.*, 2014). One of these species, *Barbarea vulgaris*, has diverged into two different 'plant types' that differ in saponin profile and herbivore resistance, as well as in other defense-related traits (Kuzina *et al.*, 2011; Hauser *et al.*, 2012; Toneatto *et al.*, 2012). One type, the G-type (Glabrous), is resistant to some specialist insect herbivores, including the devastating agricultural pest diamondback moth (*Plutella xylostella*) and the flea beetle (*Phyllotreta nemorum*). The other type, the P-type (Pubescent), is susceptible to these insects (Agerbirk *et al.*, 2003; Kuzina *et al.*, 2011). Due to the interesting defense properties of *Barbarea*, including an unusual glucosinolate profile, it has become a model for evolution and ecology of plant defense compounds (Augustin *et al.*, 2011; Byrne *et al.*, 2017).

Insect resistance in *B. vulgaris* correlates in a concentration-dependent manner with four triterpenoid saponins: the cellobiosides of hederagenin, oleanolic acid, gypsogenin, and 4-epi-hederagenin (Shinoda *et al.*, 2002; Agerbirk *et al.*, 2003; Kuzina *et al.*, 2009; Fig. 1a). These four saponins are constitutively present and can be induced to elevated levels upon insect and pathogen attack (van Mölken *et al.*, 2014). Among these

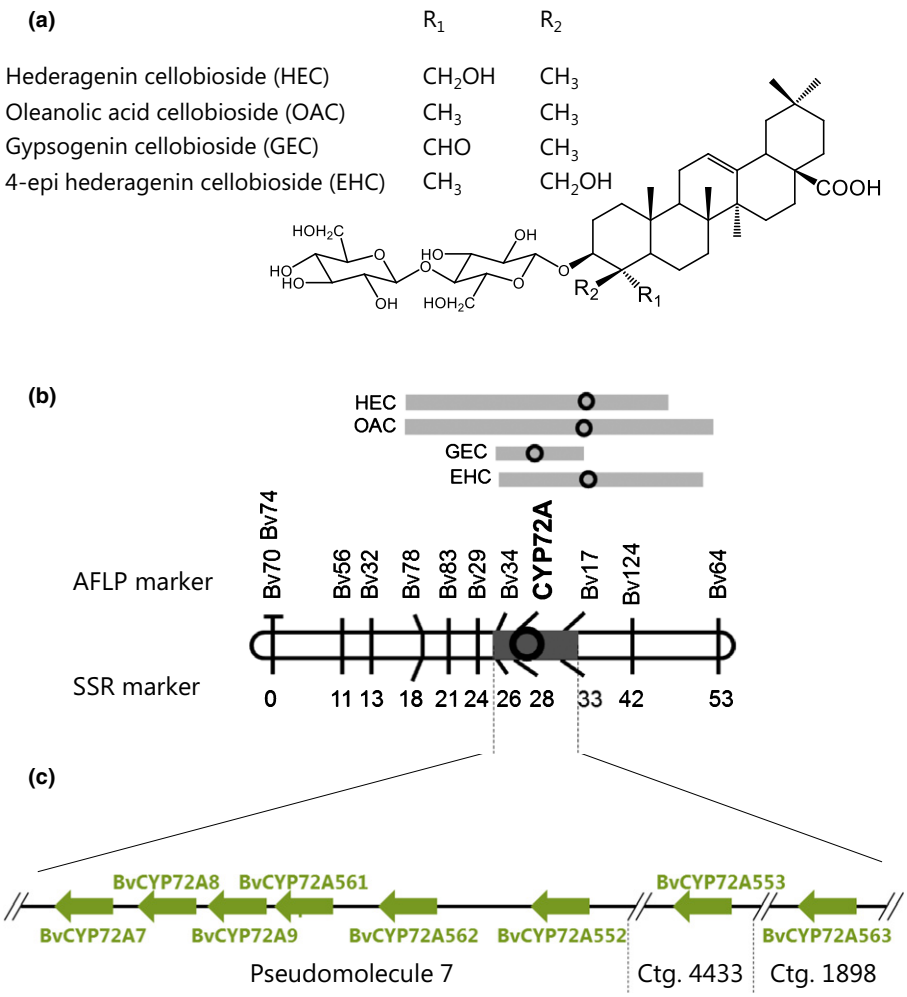

Fig. 1 Map-based cloning of *BvCYP72As*. (a) Chemical structures of the four known G-type saponins that correlate with resistance to flea beetles. (b) Location of *BvCYP72As* in the genetic map. Quantitative trait loci (QTLs) for the four G-type saponins correlating with flea beetle resistance are indicated by light-grey bars above the map, a QTL for flea beetle resistance by a dark-grey bar, and amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers on the top and bottom of the linkage group, respectively. The position of the maximal limit of detection (LOD) score within the confidence interval of the QTLs is indicated with a circle. (c) Eight candidate *CYP72A* genes identified in the QTL, of which six could be localised to pseudomolecule 7, and two on contig (ctg.) 4433 and 1898, respectively.

saponins, hederagenin cellobioside is especially toxic or deterrent, as indicated by feeding assays with purified compounds painted on radish leaf discs (Renwick, 2002; Badenes-Perez *et al.*, 2014). In comparison, oleanolic acid cellobioside, which differs from hederagenin cellobioside only by the lack of a C-23 hydroxyl group, has much less effect (Nielsen *et al.*, 2010).

Knowledge of genes and enzymes involved in the biosynthesis of saponins is incomplete. According to a proposed model, the precursor 2,3-oxidosqualene is cyclised to a limited number of core structures, which are subsequently decorated with functional groups and, finally, sugars are added (Seki *et al.*, 2015). Oxidosqualene cyclases (OSCs) catalyse the first committed step in the pathway forming the core structures, to which functional oxygenated groups are added by cytochromes P450 (P450s), and sugars by UDP-glycosyltransferases (UGTs; Seki *et al.*, 2015).

Several studies have made significant progress in uncovering genes and enzymes involved in saponin biosynthesis and the associated flea beetle resistance in *B. vulgaris*. Two OSCs, LUP5 and LUP2, have been identified and characterised from G-type and P-type plants. LUP5 is preferentially expressed in G-type (flea beetle resistant) plants and produces mainly β-amyrin as well as some α-amyrin and lupeol. β-amyrin is the backbone of the four known resistance-conferring oleane-type saponins. LUP2 is preferentially expressed in P-type (insect herbivore susceptible) plants and produces lupeol, the backbone of lupeol-type saponins, which predominate in the P-type (Khakimov *et al.*, 2015). As P-type plants are insect herbivore susceptible, we do not anticipate that lupeol-based saponins are involved in herbivore insect resistance in *B. vulgaris*. Other OSCs with similar function have been reported from licorice, ginseng, *Lotus*, *Arabidopsis thaliana*, etc. (Kushiro *et al.*, 1998; Hayashi *et al.*, 2001; Husselstein-Muller *et al.*, 2001; Sawai *et al.*, 2006). The P450s CYP716A80 and CYP716A81 from the G-type and P-type, respectively, are preferentially expressed in the G-type and oxidises β-amyrin at the C-28 position to oleanolic acid, the precursor of hederagenin (Khakimov *et al.*, 2015). Several UDP-glycosyltransferases from the UGT73C subfamily in *B. vulgaris* have been isolated (Augustin *et al.*, 2012; Erthmann *et al.*, 2018). Among these, UGT73C11 and its orthologues UCT73C10 have evolved to specifically catalyse 3-*O*-glucosylation of sapogenins (triterpenoid saponin backbones), including oleanolic acid and hederagenin.

In the *B. vulgaris* genome, two quantitative trait loci (QTL) for the four known oleanolic acid-based saponins were found to colocalise with a QTL for flea beetle resistance (Khakimov *et al.*,

2015; Byrne *et al.*, 2017). The first QTL also colocalises with LUP5, this finding is in agreement with the idea that β-amyrin-derived saponins confer resistance to flea beetles, while LUP2, catalyzing lupeol-based saponins, is not in the QTL. The other QTL is syntenic to a region on *A. thaliana* chromosome 3 containing eight highly similar *CYP72As* (Kuzina *et al.*, 2011). Several members of the CYP72A subfamily from *Glycyrrhiza* and other legumes are involved in oleanolic acid-derived saponin biosynthesis, by oxidizing at C2, C21, C22 and/or C30 positions (Seki *et al.*, 2011; Fukushima *et al.*, 2013; Biazzi *et al.*, 2015). These findings, and their location in a QTL for flea beetle resistance, suggest that CYP72As from *B. vulgaris* may be involved in the biosynthesis of hederagenin (C-23 oxidised oleanolic acid) and derived saponins, and therefore for defense against insect herbivores.

In this study, we identified *CYP72A552* colocalising with a QTL region involved in flea beetle resistance, and showed that it plays an important role in activating deterrence against the devastating agricultural pest diamondback moth and the tobacco specialist tobacco hornworm. *CYP72A552* in the saponin biosynthesis has evolved through gene duplication, most likely at some time after the origin of the *Barbarea* genus, and is still under selection.

Materials and Methods

Gene mining and cloning of CYP72A candidate genes from *B. vulgaris*

Gene mining was performed on the available draft genome of *B. vulgaris* (Byrne *et al.*, 2017): *Barbarea vulgaris* Genome Database, http://plen.ku.dk/Barbarea. Putative P450 candidates within the QTLs for flea beetles resistance were searched using the CLC Workbench (Qiagen). Eight CYP72A candidate genes were identified in the QTL region, for which PCR primer pairs (Supporting Information Table S1) were designed either from genome DNA or cDNA sequences. The full-length coding sequence of all eight candidate CYP72A genes was verified with Sanger sequencing and their sequences deposited into the NCBI database with accession numbers from MH252567 to MH252574.

Transient expression in *Nicotiana benthamiana*

For transient expression in tobacco (*Nicotiana benthamiana*), the selected genes were amplified by PCR from *B. vulgaris* cDNA and cloned into the pEAQ-HT-DEST expression vector (Sainsbury *et al.*, 2009), using USER technology (Nour-Eldin *et al.*, 2010). *Agrobacterium tumefaciens* infiltration (agro-infiltration) for transient expression in tobacco leaves was performed as described by Khakimov *et al.* (2015). After infiltration, plants were grown for another 4.5 d before harvesting for analysis. All constructs were verified by sequencing.

Expression in yeast

For expression analysis in yeast (*Saccharomyces cerevisiae*), strain INVSc1 (ThermoFisher Scientific, Waltham, MA, USA), *CYP72A552* was cloned into the pYeDP60 vector using USER cloning technology (Nour-Eldin *et al.*, 2010). pYES3-ADH-OSC1 (Seki *et al.*, 2008) and pELC-CYP716A80 was cotransformed with pYeDP60-CYP72A552 constructs, using the Frozen-EX Yeast Transformation II™ kit (Zymo Research, Freiburg im Breisgau, Germany). Yeast cells were cultured in synthetic complete medium without uracil, tryptophan or leucine medium containing 2% glucose at 30°C for 24 h. The cells were then collected and resuspended in synthetic complete medium containing 2% galactose, and cultured at 30°C for 2 d.

Yeast *in vitro* microsome assay

Yeast microsome assays were performed as described by Liu *et al.* (2014). Here, 100 µl isolated microsomal, 1 µl substrate (50 µM), 100 µl NADPH, 20 µl potassium buffer (1 M, pH 7.5), and 278 µl water were mixed and incubated for 1 h at 25°C with shaking (200 rpm), and the mixture was extracted with 1.5 ml ethyl acetate two times. The combined ethyl acetate extracts were dried with a ScanVac speed vacuum (Labogene, Lillerød, Denmark) before GC-MS analysis.

Metabolite analysis by GC-MS and LC-MS

Tobacco leaf discs were ground to fine powder under liquid N_2 with mortar and pestle. 100 mg of powder was used for either GC-MS or LC-MS analysis (Methods S1). LC-MS data were processed using MetAlign version 4.0 (www.metAlign.nl) as described in Yang *et al.* (2011) with modifications. MetAlign allowed baseline correction, noise elimination and subsequent spectral data alignment (De Vos *et al.*, 2007). The processing parameters were set to analyse from scan numbers 31-3827 at the retention time of 0.3–32.0 min. The output from MetAlign (signal intensities of each variable) was subjected to analysis of variance (ANOVA) to identify metabolites that differed between C1 and C2 with the cutoff value of two-fold change. Metabolite identification at levels 1 and 2, according to the Metabolomics Standards Initiatives (http://metabolomicssociety.org/), were assisted by molecular formula, generated from accurate mass spectrometry data (± 5 ppm), MS/MS fragmentation pattern, and the use of authentic standards.

GC-MS data were processed using DataAnalysis software (v.4.3; Bruker, Bremen, Germany). Intensities of saponins aglycones, including β-amyrin, oleanolic acid, hederagenin, gypsogenin, and gypsogenic acid, were extracted using characteristic *m/z* ions including 129, 203, 218, 320 and 471.

Nonchoice insect feeding assays

Tobacco hornworm (*Manduca sexta*) and diamondback moth (*Plutella xylostella*) eggs were stored in insect cages in a glasshouse at 26°C : 24°C, 16 h : 8 h light : dark cycle, and in a growth room 20°C : 20°C, 16 h : 8 h light : dark cycle, respectively, until the larvae hatched. Freshly hatched neonates were placed on tobacco or broccoli (*Brassica oleracea*) leaves overnight for acclimation. First instar larvae of tobacco

hornworm and third instar larvae of diamondback moth were starved for 4–5 h before used for feeding assays. Nonchoice assays were performed as described by Kumar *et al.* (2012) and Augustin *et al.* (2012) with modifications. Briefly, leaf discs (1.57 cm^2) from either tobacco agro-infiltrated with saponin biosynthetic genes or tobacco and broccoli applied with pure monoglucosides were placed into 24-well plates (15.6 mm well diameter), one leaf disc per well, with three layers of filter paper (Whatman, Little Chalfont, Buckinghamshire, UK) beneath them. Next, 60 µl of water was added onto the filter papers to secure humidity. One larva was placed on each leaf disc and fed for 24–52 h; 20−30 replicates (larvae) were used for each treatment. Consumed leaf area was measured with a portable area meter (LI-COR Inc., Lincoln, NE, USA) after 24 h and larvae survivorship was assessed after 52 h.

In vitro synthesis and nuclear magnetic resonance analysis

Monoglucosides of 3-*O*-oleanolic acid, 3-*O*-hederagenin, and 3-*O*-gypsogenic acid were synthesised by incubating the corresponding aglycone and UDP-glucose with purified UGT73C11 enzymes, as described by Augustin *et al.* (2012). The structures of the *in vitro*-produced monoglucosides were verified with either in-house standards or nuclear magnetic resonance (NMR; Methods S2, S3).

Sequence evolution analysis

Nucleotide sequences of CYP72As from *B. vulgaris* and *A. thaliana* were aligned on the amino acid level based on codons using MUSCLE and used to construct a maximum-likelihood bootstrapped phylogenetic tree using the JTT matrix-based model (Jones *et al.*, 1992). All *A. thaliana* P450 sequences were obtained from the website for Arabidopsis cytochrome P450, cytochrome *b5*, P450 reductase, β-glucosidase, and glycosyltransferases (http://www.p450.kvl.dk/; Paquette *et al.*, 2000). The tree with the highest log–likelihood (−6728.4268) was selected. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the highest log–likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories (+G, parameter = 1.1865)). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. The coding data were translated assuming a standard genetic code table. There were 522 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). Nonsynonymous to synonymous substitution rate ratios (Yang, 2007) were calculated for codon-based nucleotide alignments with the program 'codeml' from the PAML package and PAMLX (Xu & Yang, 2013). All site-models (NSsites) were tested with model 0 and branch models 1 : b (Hall, 2011).

Results

QTL map and genome-based cloning of *BvCYP72A*s

A quantitative trait locus (QTL) for oleanolic acid-derived saponins was previously found to colocalise with a QTL for flea beetle resistance in *B. vulgaris* (Kuzina *et al.*, 2009). Through fine mapping of this QTL region, a tandem repeat of six *CYP72A*s was identified (Fig. 1b) that colocalises to a region on *A. thaliana* chromosome 3 containing a tandem repeat of eight *CYP72A*s. The remaining two *CYP72A*s in the QTL were placed in *B. vulgaris* contig 4433 and contig 1898. Nucleotide-based sequence alignment of the *CYP72A*s showed that four of the *BvCYP72A* candidate genes (*BvCYP72A552*, *BvCYP72A553*, *BvCYP72A562* and *BvCYP72A563*) shared high nucleotide sequence identity (from 87% to 93%) which may partially explain why two *CYP72A*s (*BvCYP72A553* and *BvCYP72A563*) were not assembled in the same position as the rest, as assembling algorithms do not facilitate assembly of such highly repetitive regions (Byrne *et al.*, 2017; Fig. S1). Based on their phylogenetic clustering with *CYP72A*s, three of these were named *BvCYP72A7*, *BvCYP72A8* and *BvCYP72A9*, in agreement with the official P450 nomenclature system (http://drnelson.uthsc.edu/cytochromeP450.html). These three *CYP72As* are considered orthologues to *A. thaliana CYP72A7*, *CYP72A8* and *CYP72A9* and accordingly have received the same names as their *A. thaliana* orthologues. Of the remaining *BvCYP72A*s, one was identified as a pseudogene, named *CYP72A561*p, and the remaining were named *CYP72A552*, *CYP72A553*, *CYP72A562* and *CYP72A563*, respectively.

CYP72A552 oxidises oleanolic acid to hederagenin

To determine which of the *BvCYP72A* candidates is involved in the biosynthesis of oleanolic acid-derived saponins, each of the seven candidate genes were transiently expressed in tobacco (*Nicotiana benthamiana*) leaves together with two previously characterised upstream genes in the saponin biosynthesis pathway from *B. vulgaris*, *BvLUP5* and *BvCYP716A80*, to provide substrates for the CYP72As. We have previously shown that when these two genes are transiently coexpressed in tobacco leaves, oleanolic acid is the major product (Khakimov *et al.*, 2015). When *CYP72A552* was coexpressed with *BvLUP5* and *BvCYP716A80*, two major peaks and one minor could be detected (Fig. 2a); this was not found for any of the other *CYP72A* gene combinations. The two major peaks were identified as hederagenin and gypsogenic acid respectively, and the minor peak as gypsogenin, based on authentic standards. Hederagenin, gypsogenin and gypsogenic acid are the corresponding alcohol, aldehyde and carboxylic acid derivative of oleanolic acid when oxidised at the C-23 position, demonstrating that CYP72A552 oxidises at this position.

To further verify its catalytic function, *CYP72A552* was expressed in yeast cells to perform *in vitro* microsome enzyme assays. When oleanolic acid was administrated as the substrate to

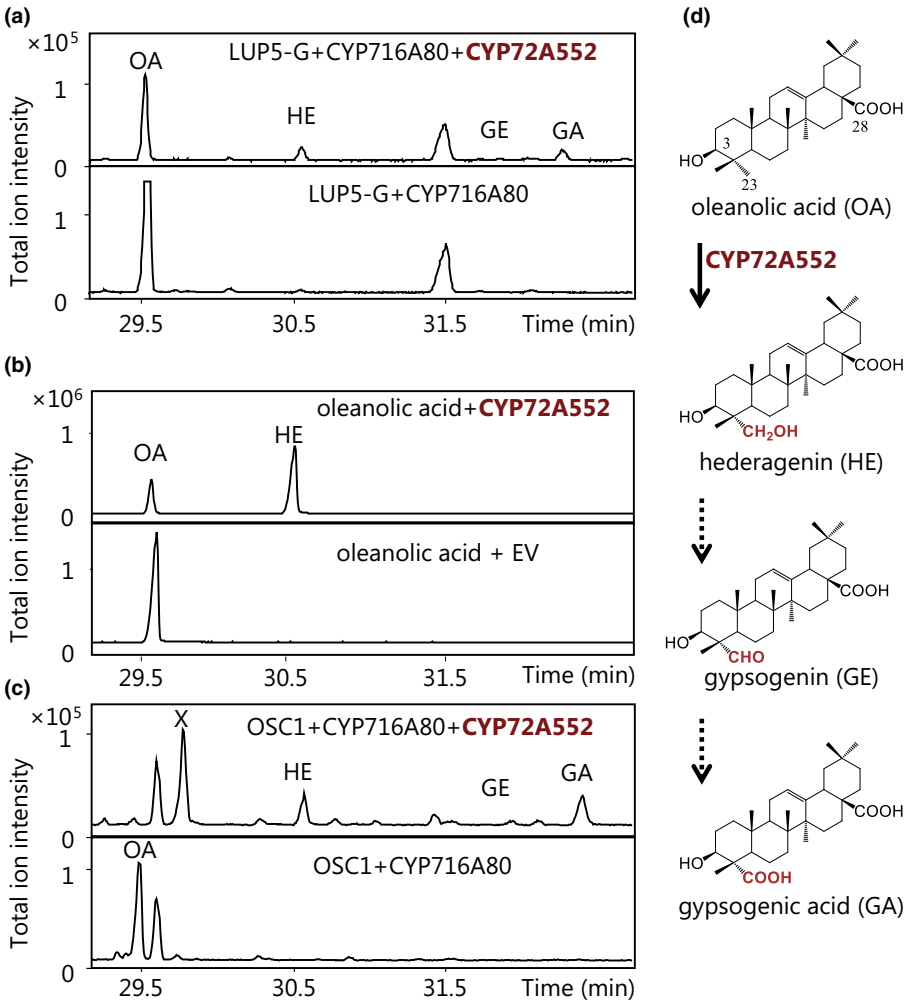

Fig. 2 CYP72A552 oxidises oleanolic acid at C-23 to hederagenin. (a) Total ion chromatogram (TIC) GC-MS profiles of tobacco (*Nicotiana benthamiana*) transiently expressing *LUP5*, *CYP716A80* and *CYP72A552*. (b) TIC GC-MS profile of yeast (*Saccharomyces cerevisiae*) microsome harbouring CYP72A552 and administrated with oleanolic acid. EV, yeast microsome of empty vector. (c) TIC GC-MS profiles of yeast expressing *LjOSC1* (*LUP5* homologue), *CYP716A80* and *CYP72A552*. Peak X is putatively identified as 23-hydroxy erythrodiol based on MS fragmentation patterns. (d) Enzymatic hydroxylation of oleanolic acid (OA) by CYP72A552; dashed arrow indicates conversion carried out by endogenous yeast cell or tobacco leaf enzymes.

yeast microsomes harbouring CYP72A552, only hederagenin was detected (Fig. 2b). We also expressed *BvCYP72A552* in yeast cells together with *LjOSC1* from *Lotus japonicus* (which produces mainly β-amyrin), and *BvCYP716A80*. This resulted in four new peaks (Fig. 2c): hederagenin, gypsogenin, gypsogenic acid, and 23-hydroxy erythrodiol, the latter putatively identified from the mass spectrum and fragmentation pattern (see Fig. S2 for the mass spectrum of each compound). To rule out if the gypsogenin and gypsogenic acid identified from the yeast and tobacco were converted by CYP72A552, hederagenin was administrated as substrate to the microsomes harboring CYP72A552. Hederagenin was not further oxidised by CYP72A552, demonstrating that CYP72A552 only oxidises oleanolic acid to hederagenin (Fig. S3). Therefore, CYP72A552 oxidises oleanolic acid to hederagenin, which is further oxidised to gypsogenin and gypsogenic acid probably by endogenous enzymes in both yeast cells and tobacco plants. When CYP72A552 was incubated with betulinic acid, the corresponding acid of lupeol oxidised at C-28, catalysis could not be detected (data not shown). In conclusion, of the eight CYP72As in the QTL for flea beetle resistance, only CYP72A552 oxidises oleanolic acid at the C-23 position to hederagenin (Fig. 2d).

Hederagenin-based saponins are a major deterrent to both diamondback moth and tobacco hornworm

Barbarea vulgaris plants accumulate both mono- and bidesmocidic saponins, with between two to five sugars attached to the C-3 or C-28 positions; the corresponding monoglycosides are usually not present (Khakimov *et al.*, 2016). The sapogenins (saponin aglycons) are not active against crucifer herbivores, but only become deterrent when they are 3-*O*-glucosylated (Nielsen *et al.*, 2010; Augustin *et al.*, 2012). The 3-*O*-monoglucosides are as active as the corresponding diglucosides and used here to determine the deterrent activity of C-23 oxidised saponins (Augustin *et al.*, 2012). Monoglucosides of oleanolic acid (OA), hederagenin (OA alcohol) and gypsogenic acid (OA acid) were synthesised *in vitro* (yielding 18.0, 15.6 and 13.0 mg of each, respectively), purified and structurally verified with in-house standards or NMR (Figs 3a, S4–S6; Tables S2, S3), and tested for antifeedant activity against the crucifer specialist herbivore diamondback moth (*P. xylostella*). The monoglucosides were applied to broccoli (*Brassica oleracea*) leaf discs at concentrations of 25 and 100 nmol cm^{-2}, and presented to diamondback moth larvae in nonchoice assays. These concentrations where chosen to

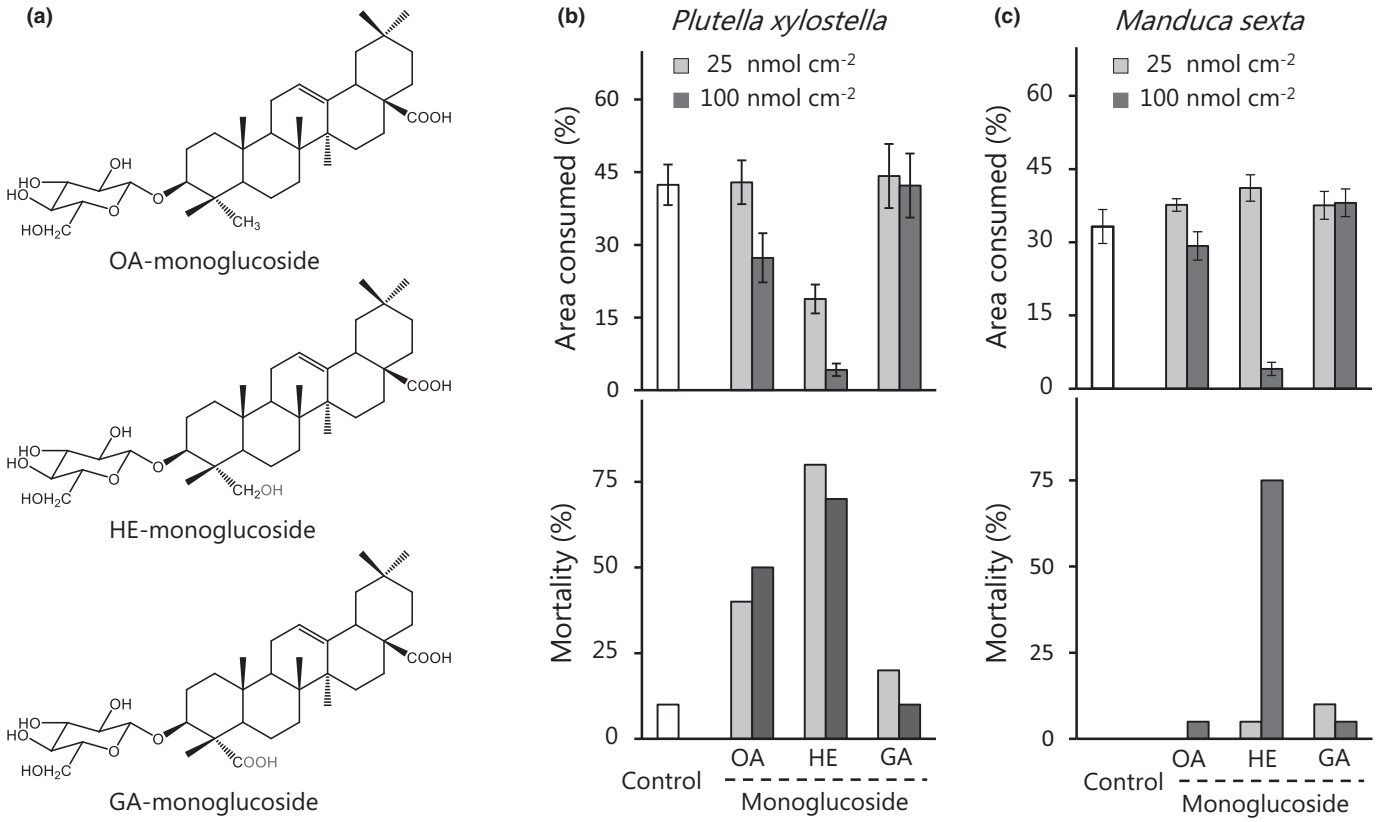

Fig. 3 Antifeedant activity of saponin monoglucosides. (a) Structures of *in vitro* synthesised oleanolic acid 3-*O*-monoglucoside (OA monoglucoside), hederagenin 3-*O*-monoglucoside (HE monoglucoside), and gypsogenic acid 3-*O*-monoglucoside (GA monoglucoside). Consumption and mortality (after 24 or 52 h, respectively) of (b) *Plutella xylostella* and (c) *Manduca sexta* larvae on leaf discs applied with OA-, HE- and GA-monoglucosides. Error bars represent the variation among three independent experiments (± SE).

reflect natural concentrations of hederagenin cellobioside in fresh leaves of *B. vulgaris*, which is *c.* 94 nmol cm^{-2}. A concentration of 100 nmol cm^{-2} hederagenin 3-*O*-monoglucoside reduced larval feeding by 90% compared with the controls, and is a four-fold stronger feeding reduction compared with oleanolic acid 3-*O*-monoglucoside (Fig. 3b). At the concentration of 25 nmol cm^{-2}, only hederagenin 3-*O*-monoglucoside reduced larval feeding, which is in agreement with the expected concentration dependency. Even more strikingly, hederagenin 3-*O*-monoglucoside caused up to 75% larval mortality, which is two-fold higher than that of oleanolic acid 3-*O*-monoglucoside or gypsogenic acid 3-*O*-monoglucoside. In conclusion, hederagenin 3-*O*-monoglucoside is a much stronger defense-compound against diamondback moth than oleanolic acid 3-*O*-monoglucoside or gypsogenic acid 3-*O*-monoglucoside, and substantiates the role of C-23 hydroxylation in saponin-mediated defense.

To develop a fast and efficient test of antifeedant effect of plant defense metabolites without the need of laborious production of pure compounds, we produced the *B. vulgaris* 3-*O*-monoglucosides in tobacco through transient gene expression and tested their antifeedant effects on tobacco hornworm (*Manduca sexta*) larvae, an important solanaceous 'specialist' herbivore. Transient gene expression is not possible in *B. vulgaris*, and diamondback moth cannot feed on tobacco. Before our tests in tobacco, we needed to verify that tobacco hornworm is indeed affected by *B. vulgaris* saponins. Tobacco leaf discs were painted with the three *in vitro* synthesised monoglucosides at the same concentrations as above. At a concentration of 100 nmol cm^{-2}, hederagenin 3-*O*-monoglucoside reduced larval feeding by 88% compared with controls, which is a seven-fold stronger feeding reduction than oleanolic acid 3-*O*-monoglucoside. Further, hederagenin 3-*O*-monoglucoside caused 75% larval mortality compared with < 5% for the other two monoglucosides (Fig. 3). At a concentration of 25 nmol cm^{-2}, none of the monoglucosides affected consumption, confirming the concentration dependency of saponins for defense. In conclusion, hederagenin 3-*O*-monoglucoside is a major deterrent or toxin against both diamond back moth and tobacco hornworm. This shows that the test system with tobacco leaf discs and tobacco hornworm larvae can be used for structure–activity relationship studies of transiently expressed saponins in tobacco.

BvCYP72A552 mediates hederagenin-based herbivore defense

Genetic modification, mutant collections and overexpression lines are not yet available in *B. vulgaris*. Therefore, to further analyse the role of CYP72A552 in saponin defense against insect herbivores, transient expression in tobacco leaves was pursued: the saponin biosynthesis pathway was first transiently

(a)

	Gene combination
C1	LUP5+CYP716A80+UGT73C11
C2	LUP5+CYP716A80+**CYP72A552**+UGT73C11
Control (CK)	Empty vector

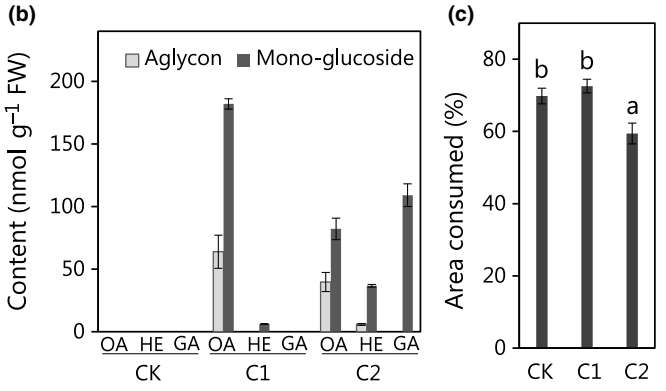

Fig. 4 BvCYP72A552 is involved in herbivore defense. (a) Combinations of saponin biosynthetic genes transiently expressed in tobacco plant (*Nicotiana benthamiana*). (b) Content of aglycons and monoglucosides of oleanolic acid (OA), hederagenin (HE) and gypsogenic acid (GA) in tobacco leaf discs transiently expressing gene combinations C1, C2 or empty vector controls (CK). (c) Consumption by *Manduca sexta* larvae of tobacco leaf discs transiently expressing C1 or C2 gene combinations compared with the empty vector control. Error bars represent the variation among three independent experiments (± SE).

reconstructed in tobacco leaves, which were subsequently presented to tobacco hornworm in nonchoice bioassays as described above. Two gene combinations, differing with respect to inclusion of *CYP72A552*, were expressed in tobacco leaves by agroinfiltration: *LUP5+CYP716A80+UGT73C11* (C1) and *LUP5+CYP716A80+CYP72A552+UGT73C11* (C2) (Fig. 4a). To verify that the pathway was properly reconstituted, the major monoglucosides, as well as sapogenins, were identified in leaf extracts and quantified by LC-MS and GC-MS, respectively (Fig. 4b). When CYP72A552 (C2) was coexpressed, hederagenin 3-*O*-monoglucoside accumulated to 36.6 nmol g^{-1} FW while concentrations of oleanolic acid 3-*O*-monoglucoside decreased to 82.1 nmol g^{-1} FW compared with in the C1 gene combination. This 55% decrease of oleanolic acid 3-O-monoglucoside compared with C1 reflects that CYP72A552 metabolises oleanolic acid to hederagenin, depleting the amount of oleanolic acid available for glucosylation. Gypsogenic acid 3-*O*-monoglucoside accumulated to 109 nmol g^{-1} FW in C2, while gypsogenic acid 3-*O*-monoglucoside was not detected in C1. As for saponin aglycons, the concentration of oleanolic acid decreased by 38% in C2 compared with C1, whereas hederagenin accumulated to 5.8 nmol g^{-1} FW in C2 compared with trace amount in C1. These results suggest that both gene combinations performed as expected in tobacco plants and that hederagenin 3-*O*-monoglucoside only accumulated when *CYP72A552* was coexpressed. The results also confirmed the presence of enzyme systems in tobacco that metabolise hederagenin further to gypsogenic acid (Fig. 2a), leading to an accumulation of the nontoxic gypsogenic acid 3-*O*-monoglucoside at the expense of toxic hederagenin 3-*O*-monoglucoside.

When tobacco leaves with the reconstituted saponin biosynthesis pathway were presented to tobacco hornworm larvae, leaf consumption was significantly reduced for C2 compared with the control (15% reduction, $P < 0.01$; Fig. 4c). By contrast, consumption was not reduced for C1. This suggests that *CYP72A552* causes the antifeedant activity of the transiently produced saponins. The relatively low antifeedant activity in the transient expression system relative to when saponins were painted on leaves (above) reflects the low concentration of hederagenin 3-*O*-monoglucoside (36.6 nmol g^{-1}) that accumulate in the tobacco leaves. In comparison, *B. vulgaris* leaves contain *c.* 40-fold more hederagenin cellobioside (*c.* 1400 nmol g^{-1}) which is similar to the concentration used on the painted broccoli leaves (1458 nmol g^{-1}).

To further determine if the antifeedant effect was directly caused by saponins that accumulated when *CYP72A552* was coinfiltrated or if other metabolites might be involved, a nontargeted metabolite analysis comparing C2 to C1 was performed using LC-MS. Analysis of variance revealed that 44 mass signals, out of 8491, were significantly upregulated in C2 compared with C1. The 44 mass signals corresponded to 15 metabolites, which were further tentatively identified as saponins containing sapogenins with one or two hexoses, including hederagenin 3-*O*-monoglucoside as verified with an authentic standard (Table S4). This result further substantiates that the deterrent effect of saponins against tobacco hornworm in tobacco leaves is an effect of CYP72A552.

BvCYP72A552 evolved through gene duplication and selection

A maximum-likelihood analysis of *CYP72A* amino acid sequences from *B. vulgaris* and *A. thaliana* was performed to reconstruct their evolutionary history. The three orthologous pairs, *CYP72A7*, *CYP72A8*, and *CYP72A9*, evolved before the two species diverged (Fig. 5a, blue mark). Subsequently, an ancestral *CYP72A* duplicated and diverged independently in the two evolutionary lineages leading to *A. thaliana* and *B. vulgaris* (Fig. 5a, red marking). Nonsynonymous to synonymous substitution ratios ($\omega = dN/dS$) were calculated to indicate whether the *CYP72A*s evolved under different selection pressures. PAML analysis showed that the ω values of *CYP72A552* and *CYP72A562* are much higher than for the other *CYP72A*s (Fig. 5a), indicating they have been subject to stronger positive selection, especially in comparison with those that evolved before the split of *A. thaliana* and *B. vulgaris*.

The order of the *CYP72A*s in the tandem array corresponds well to the topology of the phylogenetic tree (Fig. 5b). Each of the three orthologous pairs of *B. vulgaris* and *A. thaliana CYP72A*s, (*CYP72A7*, *CYP72A8*, *CYP72A9*) are placed at one end of the tandem array, however, in an inverted order, whereas the three nonorthologous genes (*BvCYP72A561P*, *BvCYP72A562*, and *BvCYP72A552*) are placed in the other end of the tandem array. Although *B. vulgaris* is closely related to *A. thaliana*, there are

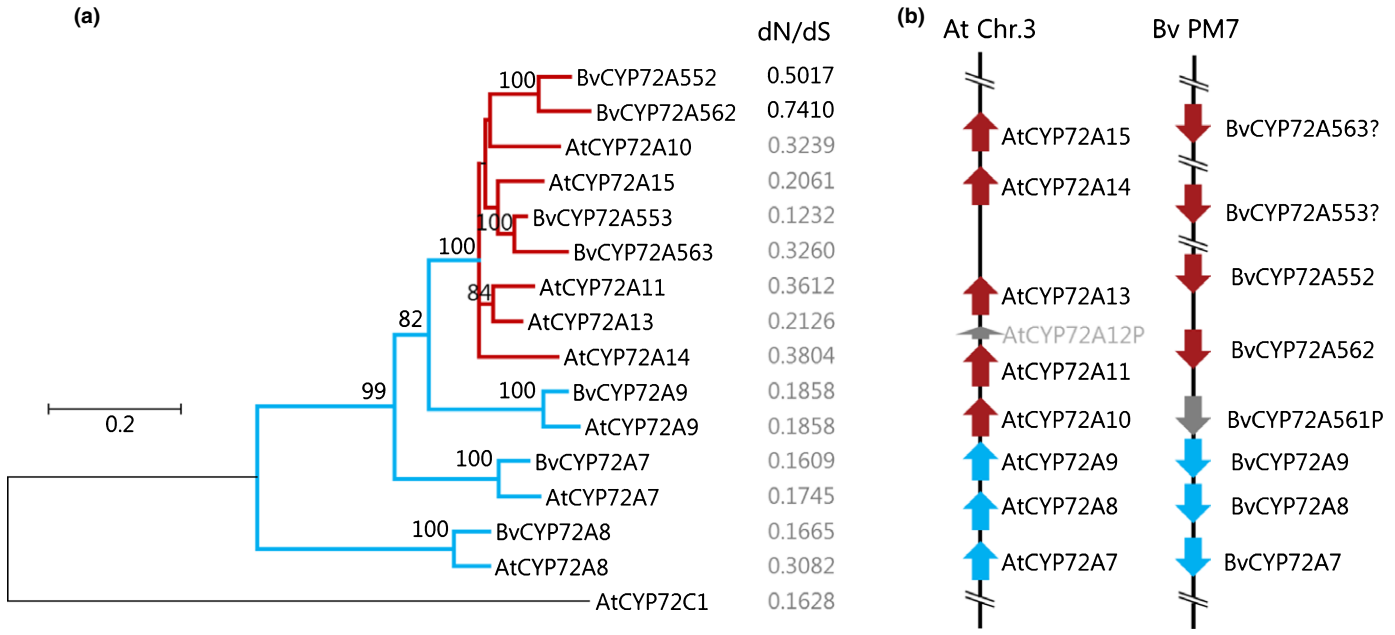

Fig. 5 Phylogenetic analysis and localisation of *CYP72A*s from *Barbarea vulgaris* and *Arabidopsis thaliana*. *B. vulgaris* sequences/genes are marked with the prefix Bv, and *A. thaliana* sequences/genes are marked with an At prefix. Phylogenetic clades and genes inferred to have evolved before the split of the *B. vulgaris* and *A. thaliana* lineages are marked with blue, those evolved after the split in red. (a) Maximum-likelihood phylogenetic tree of *CYP72A* sequences from *B. vulgaris* and *A. thaliana*, using *AtCYP72C1* as out-group, and with associated dN/dS values. Only bootstrap values above 70% are shown, and the bar indicates 0.2 substitutions per site. (b) Selection of *A. thaliana* chromosome 3 and *B. vulgaris* pseudomolecule 7 illustrating the region of the *CYP72A* subfamily in these two species. The broken line denotes an undissolved region. The genes (*BvCYP72A563* and *BvCYP72A553*) anchored to this undissolved region on pseudomolecule 7 are denoted by a question mark (?).

multiple cases of inversions and translocations (Byrne *et al.*, 2017). So, the unexpected inverted relative order of the *A. thaliana* and *B. vulgaris* CYP72As, may either be a consequence of such genome reorganisations, or that the draft genome of *B. vulgaris* is rather crude and many regions, including the two QTLs, were difficult to assemble (Byrne *et al.*, 2017). The bootstrapped tree and ω values indicated that *BvCYP72A552* evolved through recent gene duplication. In addition, due to the high sequence similarity and identity among the *CYP72A* sequences, the phylogenetic relationships between *B. vulgaris* and *A. thaliana* *CYP72A552* cannot be resolved, and is reflected by the absence of bootstrap values above 70% (Fig. 5).

In summary, our results suggest that evolution of the saponin-based plant defense in the *Barbarea* genus was mediated by gene duplications followed by specializing selection.

Discussion

Combined, *B. vulgaris* G-type and P-type plants accumulate up to 49 different saponins structures, as verified by LC-MS-NMR (Khakimov *et al.*, 2016). Some of these saponin structures are more efficient in mediating resistance to insect herbivores than others (Agerbirk *et al.*, 2003; Kuzina *et al.*, 2009; Augustin *et al.*, 2012). Four oleanolic acid-derived saponins from *Barbarea vulgaris* are known to be constitutively produced and induced upon flea beetle attack (Kuzina *et al.*, 2011; Toneatto *et al.*, 2012), and they have been shown to correlate with flea beetles resistance in two QTLs. The QTL on linkage group 1 controls expression of *LUP5*, while the other QTL is known to contain a tandem array of eight *CYP72A*s (Shinoda *et al.*, 2002; Kuzina *et al.*, 2011; Byrne *et al.*, 2017). Here we demonstrate that only one of them, *CYP72A552*, is responsible for oxidizing oleanolic acid to its corresponding alcohol, hederagenin, and that the monoglucoside of hederagenin is a major deterrent to the crucifer specialist herbivore diamondback moth and the solanaceous facultative 'specialist' tobacco hornworm. We have previously shown that the hederagenin 3-*O*-glucoside is a stronger deterrent than oleanolic acid 3-*O*-glucoside for flea beetles (Augustin *et al.*, 2012). Transient expression of *CYP72A552* in tobacco leaves, together with other known genes in the *B. vulgaris* saponin biosynthesis pathway, led to the accumulation of hederagenin-derived saponins and significantly reduced feeding by tobacco hornworm. Our results indicated that *BvCYP72A552* evolved through gene duplication after the split between the *A. thaliana* and *B. vulgaris* lineages and is under stronger positive selection than the CYP72As before the split. Collectively, *CYP72A552* amplifies the deterrent effect of the oleanolic-based saponins and therefore is involved in mediating insect herbivore defense in *B. vulgaris*. The fact that CYP72A552 is expressed in both the G-type and P-type (Fig. S7; Methods S4) and only oxidises oleanolic acid and not lupeol-based sapogenins, substantiates that it is the ability to produce β-amyrin and to glucosylate at the C-3 position that are the main evolutionary determinates for saponin-based insect herbivore resistance in *B. vulgaris*.

In previous studies by Fukushima *et al.* (2013) and Kim *et al.* (2018), *CYP72A68v2* from *Medicago truncatula* and *CYP714E19* from *Centella asiatica*, respectively, were reported to catalyse up to three consecutive hydroxylations of oleanolic acid

at the C-23 position to first hederagenin, then gypsogenin and last gypsogenic acid by one P450. In this study, we show that when *CYP72A552* was coexpressed with upstream biosynthesis genes in tobacco leaves or yeast cells hederagenin, gypsogenin, and gypsogenic acid could be detected. However, our *in vitro* enzyme assay, using isolated yeast microsomes harbouring CYP72A552, showed that CYP72A552 exclusively oxidises oleanolic acid to hederagenin. Therefore, we concluded that CYP72A552 oxidises oleanolic acid at the C-23 position only to hederagenin, and that yeast and tobacco enzymes can further oxidise hederagenin to gypsogenic acid. In agreement with this conclusion, Han *et al.* (2017) found that hederagenin was the only product of yeast microsomes harbouring *Kalopanax septemlobus* CYP72A397 when administrated with oleanolic acid. So, the gypsogenin-derived saponins found in *B. vulgaris* must be catalysed by another enzyme instead of CYP72A552.

B. vulgaris, and a few other closely related species, are the only crucifers known to produce saponins (Badenes-Perez *et al.*, 2014). None of the eight CYP72As located on *A. thaliana* chromosome 3 could oxidise oleanolic acid to hederagenin or other compounds (data not shown). We have previously shown that *B. vulgaris* has evolved a UDP-glycosyltransferase, UGT73C11, that specifically catalyses 3-*O*-glucosylation of sapogenins (Augustin *et al.*, 2012), and does not have a functional homologue in *A. thaliana*. Therefore, at least two of the genes, *UGT73C10/C11* and *CYP72A552*, involved in *Barbarea* saponin biosynthesis seem to have evolved and specialised subsequent to their divergence from the *A. thaliana* lineage. In both cases, this seems to have been mediated by gene duplication(s) of biosynthetic genes, initially shared with a common ancestor, followed by positive selection for novel function in the *Barbarea* lineage. In contrast with *LUP5* and *CYP72A552*, none of *UGT73C11*, *CYP716A*80 or *CYP716A*81 was found close to the two QTLs for flea beetle resistance in *B. vulgaris*. This finding is to be expected, as all saponins are glucosylated at C-3 (UGT73C10/C11) and most of them are carboxylated at C-28 (CYP716A80/81), and not all *Barbarea* saponins contribute to resistance. By contrast, only some *B. vulgaris* saponins are hydroxylated at C-23 (Khakimov *et al.*, 2016) and among these are the defense-conferring hederagenin-based saponins. *CYP72A552* is expressed in both G-type and P-type *B. vulgaris* (Fig. S7), and this may reflect that it is colocalised with other genes related with resistance to flea beetles. We have previously shown that CYP716s and OSCs are promiscuous enzymes (Khakimov *et al.*, 2015), both with respect to substrate and product specificity, and this may explain how *B. vulgaris* with a limited number of genes in the pathway may accumulate up to 49 different saponins (Khakimov *et al.*, 2016).

Our study provides important new insight into the structure–activity relationships of saponins, by showing that oxidative modification of the triterpenoid sapogenin may strongly increase or decrease bioactivity. A previous study has shown that a carboxyl group attached to the C-28 of oleanolic acid is crucial for antitumor and hemolytic activity (Carelli *et al.*, 2011; Sui *et al.*, 2017). Here, we showed that adding a hydroxyl group at C-23 significantly decreased feeding and survival of diamondback moth and tobacco hornworm. These results substantiates previous results with the crucifer specialist flea beetle that originally was used to identify the QTLs (Nielsen *et al.*, 2010; Augustin *et al.*, 2012; Christensen *et al.*, 2019).

The actual mode of action of saponins is largely unknown and we can therefore only speculate why the C-23 hydroxy group is so crucial for biological activity. Preliminary analysis of structural changes through molecular energy minimisation modelling indicates that the glucose unit at the 3-*O* position of oleanolic acid monoglucoside is twisted *c.* 90° relative to the plane of the sapogenin backbone when a hydroxyl group is added at position C-23, leading to formation of hederagenin (Fig. S8). However, when a carboxyl group is introduced instead, the glucose unit is further twisted to a total of *c.* 180° bringing the glucose more or less back to being in the same plane as the sapogenin backbone. Therefore, the presence of the hydroxyl group at C-23 appears to be crucial for how the C-3 glucose is oriented towards the triterpenoid backbone, and this situation may be important for how saponins exert their toxicity at the molecular level. Saponins are generally believed to disrupt membranes by forming complexes with membrane sterols or to interact with steroidal receptors (Augustin *et al.*, 2011). Possibly, the C-23 hydroxylation by CYP72A552 induces a rotation in the saponin molecule changing its physiochemical properties, and this could possibly trigger membrane disruption, affinity to steroidal receptors, or decreases efficiency of saponin detoxifying enzymes in larval guts.

In this study, we introduced a fast and powerful system for determining structure–activity relationships by transiently expressing biosynthetic genes in tobacco leaves for saponin production and subsequent feeding assays with insects. This complements traditional feeding studies, in which the metabolites are applied to leaf surfaces, by being more physiologically realistic as the tested compounds are produced and stored within leaves. The transient system takes < 1 wk from introduction of the genes to insect bioassays, compared with months needed for stably transformed plants. Furthermore, the transient expression system allows the production of compounds that may otherwise be toxic or lethal to the plant if expressed systemically.

An unexpected drawback of the transient tobacco system is that hederagenin was further metabolised by endogenous tobacco enzymes, hampering the accumulation of hederagenin-3-*O*-monoglucoside. Therefore, hederagenin 3-*O*-monoglucoside accumulated to 36.6 nmol g^{-1} in the tobacco leaves (corresponding to 2.5 nmol cm^{-2}), whereas natural concentrations of hederagenin cellobioside in flea beetle-resistant *B. vulgaris* plants is *c.* 1400 nmol g^{-1} (corresponding to *c.* 96 nmol cm^{-2}) and concentrations used in our leaf disc assays painted with the monoglucosides were 100 nmol cm^{-2}. The relatively modest decrease in feeding on transiently saponin-expressing tobacco (15% reduction), in comparison to the feeding assays with saponin-painted leaf discs at the high concentration (90% reduction), is most likely to be due to the much lower concentration of saponins.

In summary, our study provides a key link between *CYP72A552* and saponin-mediated plant defense against insects and reveals that its function has evolved through gene duplication and selection for novel function. Our study further highlights the

evolution of chemical novelties by gene duplication, recruitment of enzymes into new functions, and the importance of chemical modification in plant defense evolution.

Acknowledgements

We thank Professors Ian Baldwin and Marcel Dicke for providing *Manduca sexta* and *Plutella xylostella* material, respectively, Professor Giovanni Battista Appendino, Alberto Minassi and Federica Pollastro for providing saponin aglycon standards, Assistant Professor Henrik Hjarvard de Fine Licht for helping with sequence evolution analysis and Professor David Nelson for naming CYP72As according to the P450 nomenclature. This work was supported by a Marie Skłodowska-Curie European Fellowship (H2020-MSCA-IF-2014, grant no. 655559 to SSAR; H2020-MSCA-IF-2016, grant no. 752437 to REMES), the Danish Council for Independent Research, Technology and Production Sciences (grant no. 1335-00151), the European Community's Seventh Framework Programme (FP7/2007-2013; grant no. 613692 to TriForC). The authors declare no conflict of interest.

Author contributions

QL, TPH, and SB designed research; QL, BK, PDC and KRM performed research; FC and CEO contributed new reagents or analytic tools; QL and BK analysed data; QL, TPH and SB wrote the paper.

ORCID

Søren Bak https://orcid.org/0000-0003-4100-115X
Pablo D. Cárdenas https://orcid.org/0000-0002-4141-2410
Federico Cozzi https://orcid.org/0000-0001-6707-8898
Thure Pavlo Hauser https://orcid.org/0000-0001-9833-9579
Bekzod Khakimov https://orcid.org/0000-0002-6580-2034
Qing Liu https://orcid.org/0000-0002-0257-0694
Carl Erik Olsen https://orcid.org/0000-0002-2275-0596

References

Agerbirk N, Olsen CE, Bibby BM, Frandsen HO, Brown LD, Nielsen JK, Renwick JAA. 2003. A saponin correlated with variable resistance of *Barbarea vulgaris* to the diamondback moth *Plutella xylostella*. *Journal of Chemical Ecology* **29**: 1417–1433.

Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, Olsen CE, Hansen EH, Kuzina V, Ekstrøm CT. 2012. UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiology* **160**: 1881–1895.

Augustin JM, Kuzina V, Andersen SB, Bak S. 2011. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**: 435–457.

Badenes-Perez FR, Reichelt M, Gershenzon J, Heckel DG. 2014. Using plant chemistry and insect preference to study the potential of *Barbarea* (Brassicaceae) as a dead-end trap crop for diamondback moth (Lepidoptera: Plutellidae). *Phytochemistry* **98**: 137–144.

Biazzi E, Carelli M, Tava A, Abbruscato P, Losini I, Avato P, Scotti C, Calderini O. 2015. CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Molecular Plant* **8**: 1493–1506.

Byrne SL, Erthmann PØ, Agerbirk N, Bak S, Hauser TP, Nagy I, Paina C, Asp T. 2017. The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. *Scientific Reports* **7**: 40728.

Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S. 2011. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* **23**: 3070–3081.

Christensen S, Enge S, Jensen KR, Müller C, Kiær LP, Agerbirk N, Heimes C, Hauser TP. 2019. Different herbivore responses to two co-occurring chemotypes of the wild crucifer *Barbarea vulgaris*. *Arthropod-Plant Interactions* **13**: 19–30.

De Geyter E, Lambert E, Geelen D, Smagghe G. 2007. Novel advances with plant saponins as natural insecticides to control pest insects. *Pest Technology* **1**: 96–105.

De Geyter E, Swevers L, Soin T, Geelen D, Smagghe G. 2012. Saponins do not affect the ecdysteroid receptor complex but cause membrane permeation in insect culture cell lines. *Journal of Insect Physiology* **58**: 18–23.

De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2**: 778.

Erthmann PØ, Agerbirk N, Bak S. 2018. A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono- and bisdesmosidic saponins. *Plant Molecular Biology* **97**: 37–55.

Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K, Muranaka T. 2013. Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant and Cell Physiology* **54**: 740–749.

Hall BG. 2011. *Phylogenetic trees made easy: a how-to manual*. Oxford, UK: Oxford University Press.

Han JY, Chun J-H, Oh SA, Park S-B, Hwang H-S, Lee H, Choi YE. 2017. Transcriptomic analysis of *Kalopanax septemlobus* and characterization of *KsBAS*, *CYP716A94*, and *CYP72A397* genes involved in hederagenin saponin biosynthesis. *Plant and Cell Physiology* **59**: 319–330.

Hauser TP, Toneatto F, Nielsen JK. 2012. Genetic and geographic structure of an insect resistant and a susceptible type of *Barbarea vulgaris* in western Europe. *Evolutionary Ecology* **26**: 611–624.

Hayashi H, Huang P, Kirakosyan A, Inoue K, Hiraoka N, Ikeshiro Y, Kushiro T, Shibuya M, Ebizuka Y. 2001. Cloning and characterization of a cDNA encoding beta-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biological & Pharmaceutical Bulletin* **24**: 912–916.

Husselstein-Muller T, Schaller H, Benveniste P. 2001. Molecular cloning and expression in yeast of 2,3-oxidosqualene–triterpenoid cyclases from *Arabidopsis thaliana*. *Plant Molecular Biology* **45**: 75–92.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**: 275–282.

Khakimov B, Kuzina V, Erthmann PØ, Fukushima EO, Augustin JM, Olsen CE, Scholtalbers J, Volpin H, Andersen SB, Hauser TP. 2015. Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. *The Plant Journal* **84**: 478–490.

Khakimov B, Tseng LH, Godejohann M, Bak S, Engelsen SB. 2016. Screening for triterpenoid saponins in plants using hyphenated analytical platforms. *Molecules* **21**: 1614.

Kim OT, Um Y, Jin ML, Kim JU, Hegebarth D, Busta L, Racovita RC, Jetter R. 2018. A novel multifunctional C-23 oxidase, CYP714E19, is involved in asiaticoside biosynthesis. *Plant and Cell Physiology* **59**: 1200–1213.

Kumar P, Pandit SS, Baldwin IT. 2012. Tobacco rattle virus vector: a rapid and transient means of silencing *Manduca sexta* genes by plant mediated RNA interference. *PLoS ONE* **7**: e31347.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870–1874.

Kushiro T, Shibuya M, Ebizuka Y. 1998. β-Amyrin synthase: cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *European Journal of Biochemistry* **256**: 238–244.

Kuzina V, Ekstrøm CT, Andersen SB, Nielsen JK, Olsen CE, Bak S. 2009. Identification of defense compounds in *Barbarea vulgaris* against the herbivore

Phyllotreta nemorum by an ecometabolomic approach. *Plant Physiology* **151**: 1977–1990.

Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB. 2011. *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. *Phytochemistry* **72**: 188–198.

Liu Q, Manzano D, Tanić N, Pesic M, Bankovic J, Pateraki I, Ricard L, Ferrer A, de Vos R, van de Krol S *et al.* 2014. Elucidation and *in planta* reconstitution of the parthenolide biosynthetic pathway. *Metabolic Engineering* **23**: 145–153.

Nielsen JK, Nagao T, Okabe H, Shinoda T. 2010. Resistance in the plant, *Barbarea vulgaris*, and counter-adaptations in flea beetles mediated by saponins. *Journal of Chemical Ecology* **36**: 277–285.

Nour-Eldin H, Geu-Flores F, Halkier B. 2010. USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. In: Fett-Neto AG, ed. *Plant secondary metabolism engineering*. Totowa, NJ, USA: Humana Press, 185–200.

Paquette SM, Bak S, Feyereisen R. 2000. Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA and Cell Biology* **19**: 307–317.

Renwick JAA. 2002. The chemical world of crucivores: lures, treats and traps. *Entomologia Experimentalis et Applicata* **104**: 35–42.

Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* **7**: 682–693.

Sawai S, Shindo T, Sato S, Kaneko T, Tabata S, S-i Ayabe, Aoki T. 2006. Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Science* **170**: 247–257.

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T. 2008. Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences, USA* **105**: 14204–14209.

Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K. 2011. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* **23**: 4112–4123.

Seki H, Tamura K, Muranaka T. 2015. P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant and Cell Physiology* **56**: 1463–1471.

Shinoda T, Nagao T, Nakayama M, Serizawa H, Koshioka M, Okabe H, Kawai A. 2002. Identification of a triterpenoid saponin from a crucifer, *Barbarea vulgaris*, as a feeding deterrent to the diamondback moth, *Plutella xylostella*. *Journal of Chemical Ecology* **28**: 587–599.

Sui J-J, Zhou W-H, Liu D-Y, Li M-Q, Sun J-S. 2017. Highly efficient synthesis of bioactive oleanane-type saponins. *Carbohydrate Research* **452**: 43–46.

Toneatto F, Hauser TP, Nielsen JK, Ørgaard M. 2012. Genetic diversity and similarity in the *Barbarea vulgaris* complex (Brassicaceae). *Nordic Journal of Botany* **30**: 506–512.

van Mölken T, Kuzina V, Munk KR, Olsen CE, Sundelin T, van Dam NM, Hauser TP. 2014. Consequences of combined herbivore feeding and pathogen infection for fitness of *Barbarea vulgaris* plants. *Oecologia* **175**: 589–600.

Xu B, Yang Z. 2013. PAMLX: a graphical user interface for PAML. *Molecular Biology and Evolution* **30**: 2723–2724.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang T, Stoopen G, Yalpani N, Vervoort J, de Vos R, Voster A, Verstappen FWA, Bouwmeester HJ, Jongsma MA. 2011. Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering* **13**: 414–425.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Amino acid sequences alignment of the seven *CYP72A*s from *B. vulgaris* possible involved in saponin biosynthesis and the eight syntenic *A. thaliana CYP72*s.

Fig. S2 Mass spectrum of oleanolic acid, hederagenin, gypsogenin, gypsogenic acid and compound X.

Fig. S3 GC-MS profile of yeast (*S. cerevisiae*) microsome expressing *CYP72A552* administrated with hederagenin (HE).

Fig. S4 Large-scale production of monoglucosides of oleanolic acid (OA), hederagenin (HE), and gypsogenic acid (GA).

Fig. S5 Purification of the monoglucosides of oleanolic acid (OA), hederagenin (HE), and gypsogenic acid (GA).

Fig. S6 Structure of 3-*O*-monoglucoside gypsogenic acid elucidated by NMR.

Fig. S7 Gene expression of *CYP72A552* in G- and P-type *B. vulgaris* leaves by real-time RT-PCR.

Fig. S8 3D structure of 3-*O*-oleanolic acid monoglucoside (a), 3-*O*-hederagenin monoglucoside (b), and 3-*O*-gypsogenic acid monoglucoside (c).

Methods S1 GC-MS and LC-MS analysis.

Methods S2 Large-scale production of oleanolic acid, hederagenin, and gypsogenic acid monoglucosides.

Methods S3 NMR structure elucidation of gypsogenic acid monoglucoside.

Methods S4 Gene expression analysis by real-time RT-PCR.

Table S1 Primers used in this study.

Table S2 Conditions of NMR experiments.

Table S3 1D ^{1}H and HSQC-based ^{13}C NMR data of 3-*O*-monoglucoside gypsogenic acid.

Table S4 Tentative identification of metabolites significantly changed between plants transiently expressing the two saponin biosynthetic gene combinations.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.